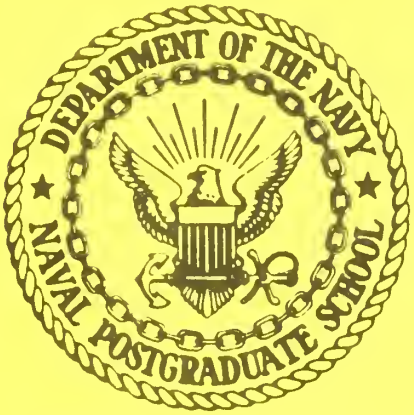
NPS55-85-020

NAVAL POSTGRADUATE SCHOOL

Monterey, California



RANDOM PARAMETER MARKOV POPULATION PROCESS MODELS AND THEIR LIKELIHOOD, BAYES, AND EMPIRICAL BAYES ANALYSIS

by

Donald P. Gaver
John P. Lehoczky

September 1985

Approved for public release; distribution unlimited.

Prepared for:
Office of Naval Research
) North Quincy Street
lington, VA 22217

NAVAL POSTGRADUATE SCHOOL
Monterey, California

Rear Admiral R. H. Shumaker
Superintendent

David A. Schrady
Provost

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NPS55-85-020	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* RANDOM PARAMETER MARKOV POPULATION PROCESS MODELS AND THEIR LIKELIHOOD, BAYES, AND EMPIRICAL BAYES ANALYSIS		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR*(s)* Donald P. Gaver John P. Lehoczky		8. CONTRACT OR GRANT NUMBER*(s)*
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, CA 93943-5100		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61153N; RR014-05-01 N0001485WR24061
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research 800 North Quincy Street Arlington, VA 22217		12. REPORT DATE September 1985 13. NUMBER OF PAGES 24
14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)*		15. SECURITY CLASS. *(of this report)* Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Markov processes, likelihood estimation, empirical Bayes

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

Markov population stochastic processes are useful in describing repairman and logistics problems, networks of queues, pharmacological processes, and manpower situations. This paper considers statistical estimation problems arising for such models. Parameter estimation of an empirical Bayes nature, with "limited shrinkage" or discrepancy tolerant" features is discussed and illustrated.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

TABLE OF CONTENTS

1 Introduction

Much of the career of Herbert Solomon has been devoted to combining the identification of important applications, the building of models for these applications and the use of data to estimate the parameters of the models or to develop new models. His activities have covered the vast territory of statistics, operations research and the social sciences. He has considered problems arising in such diverse areas as logistics, inventories and queueing, quality control and inspection, learning and human factors, packing and geometrical probability. It is in this eclectic spirit that we consider a problem area involving both probability modelling and statistical inference.

This paper presents a natural generalization and a statistical analysis for a general and important class of stochastic processes, simple Markov population processes (SMPP). This class of processes is broad enough to encompass simple queues and complex queueing networks, repair models, manpower models, labor mobility and migration phenomena, and many others. The equilibrium theory for the SMPP is well known. The reader should consult Kingman (1969) or Kelly (1979) for a thorough treatment of the probabilistic aspects of this theory. We will focus on statistical inferences associated with random parameter versions of the SMPP.

Traditional formulations of applied probability models typically emphasize only one of the many possible sources of random fluctuations that may influence system behavior. For example, extensive treatment has been given to a wide variety of queueing models for congestion at service facilities. In these models, some simple arrival or demand process confronts a given service process often presumed to have i.i.d. random variable, or possibly Markovian, character. The parameters of the component arrival and service processes are nearly always assumed to be few in number, given and fixed. Consequently, the predicted waiting times and queue lengths vary only in response to the inherent variability of the arrival and service processes if the parameters of the latter are also fixed and known. It may be argued that such internal or within variability is not the only source to be considered under many circumstances: the fixed rates may be expected to change occasionally, and additionally may well be unknown. Similarly, inventory control models typically assume that parameters of demand distributions are fixed, as do reliability-redundancy studies of failure-prone repairable systems, and the compartment models of pharmacology.

Perhaps for reasons of mathematical tractability, there has been far less attention paid to models having time-dependent parameters, where the time dependence is either deterministic or "random." Recently, however, models representing random environments or doubly stocnastic effects have appeared in the literature. Such models should be useful in system studies for describing realistic situations in which

parameter values vary widely because of weather or other natural environmental influences or because of the impact of client, patient, or opponent action, or various other exogenous influences. Thus, the traditional models should be extended by incorporating external or between variability components. Because it is often natural to suppose first that the basic parameters of the model of interest are drawn from a given (with unknown parameters) superpopulation, we refer to such models as hierarchical.

The objective of this paper is to formulate and solve problems of statistical inference for hierarchical models which arise in the context of Markov population processes. We will assume that we are given an observation of the sample path from a set of Markov population processes. Each of the processes is governed by its own set of parameters. The vectors of parameters for each of the processes are assumed to be drawn i.i.d. from a parent "superpopulation" distribution which may itself have unknown parameters. The goal is to estimate the parameters of the superpopulation, because these describe the population. Using this information, we wish to estimate the parameters of each of the Markov population processes to make inferences about a particular instance of the system, population or compartment. If positive indications are present, it will be possible to pool information from other processes to improve the estimates of one in particular, i.e., to "borrow strength" in John Tukey's words. On the other hand, we seek approaches which are "discrepancy-tolerant" or "robust," i.e., which do not unjustifiably pool data when counter-indications are presented by the data; Gaver (1985).

The general inference problem described seems to fall into the category of a standard Bayesian analysis or, more ambitiously, an empirical Bayes or Bayes-empirical Bayes analysis. The reader might consult Morris (1983) for a review of parametric empirical Bayes methods, Robbins (1983) and in much previous work, has elucidated the non-parametric Bayes approach, or Deely and Lindley (1981). Surprisingly, scant attention seems to have been paid to empirical Bayes methods for stochastic processes.

This paper is organized as follows. We introduce the notion of Markov population processes in Section 2. Section 3 gives several examples of situations in which random parameter versions of Markov population processes are important. Section 4 develops a likelihood function and Bayesian statistical inference for these processes. Section 5 briefly discusses empirical Bayes approaches. Section 6 points out important topics for future research, much of which is currently in progress.

2 Markov Population Processes

We are concerned with Markov population processes in n dimensions. Such a process $\{\underline{X}(t), t \geq 0\}$ is a continuous time Markov process with state space $S = \{(i_1, i_2, \ldots, i_n),\ i_j \in Z^+, 1 \leq j \leq n\}$. The components of the state space correspond to the occupancy of each of the n populations or colonies. The possible transitions are defined as follows. Let $T_{ij}(\underline{s}) = T_{ij}(s_1, \ldots, s_n) = (s_1, \ldots, s_{i-1}, s_i - 1, s_{i+1}, \ldots, s_j + 1, \ldots, s_n)$. The transition from state $\underline{s}$ to $T_{ij}\underline{s}$ corresponds to a migration of an individual or unit in population (colony) i to population j . For a simple Markov population process the flow rate for such transitions can depend on $\underline{s}$ only through s_i . We specialize this requirement further by requiring the flow rate to have the form $\lambda_i f_{ij}(s_i)$, $1 \leq i,j \leq n$, $i \neq j$.

We also allow transitions from the outside into one of the colonies. Let $T_{oj}(\underline{s})$ correspond to an arrival from the outside to colony j. This transition has flow rate $\lambda_o f_{oj}(s_j)$ and depends on $\underline{s}$ only through s_j. Finally, there can be departures from a colony to the outside. We define $T_{io}(\underline{s}) = (s_1, \ldots, s_i - 1, \ldots, s_n)$. The transition from $\underline{s}$ to $T_{io}(\underline{s})$ has flow rate $\lambda_i f_{io}(s_i)$.

The functions $\{f_{ij}(k),\ 0 \leq i \leq n,\ 0 \leq j \leq n\}$ can often, but not always, be thought of as known structural parameters. If one is studying a Markov population process such as a queueing network or compartmental model, these parameters are determined directly from knowledge of the structure (e.g., the compartment connectivity or the number of servers at a node and the flow of traffic among the nodes). The rate parameters $\underline{\lambda} = (\lambda_o, \lambda_1, \ldots, \lambda_n)$ are generally not known and must be estimated from data. Several comments are in order. First, a single unknown input parameter may not be sufficient, as there could be flows into different colonies whose relative magnitudes are not known. The formulation can be easily extended to allow for as many as n such parameters (one for each colony), but we do not do so here. Second, one must introduce conditions on $\underline{\lambda}$ and the structural parameters jointly to ensure that an equilibrium distribution exists. Indeed, if an equilibrium distribution were to exist, it would be of product form; see Kelly (1979). Fortunately, the existence of an equilibrium is unnecessary for our analysis. There is no need for the process to be in equilibrium or to even have an equilibrium distribution. We merely observe a part of the sample path and then develop estimates even if the process is transient. Similarly, it is usual to assume the state space is irreducible, but this is not necessary for our purposes. For this reason, we define $S = \{(i_1, \ldots, i_n),\ i_j \in Z^+,\ 1 \leq j \leq n\}$ and do not address the possible boundedness of certain components.

3 Examples of Simple Markov Population Processes

We present several examples to illustrate a few of the important applications of Markov population processes. We also describe the accompanying inference problems.

3.1 Simple Markovian Queueing Systems: M/M/C

In this example, we let n = 1. The state variable, x, corresponds to the number of customers in or awaiting service. The parameter λ_0 corresponds to the arrival rate, while λ_1 is the service rate. The structural parameters are given by $f_{01}(x) = 1$ and $f_{10}(x) = \min(x,C)$ where C is the number of servers. It is likely that either λ_0 or λ_1 or both are unknown and must be estimated from data. Not only may λ_0 be unknown, but it can also exhibit fluctuations over time, e.g., from day to day. This is especially true in service systems where demands vary. It is reasonable to gather data over relatively short periods of time during which λ_0 and λ_1 can be considered to be constant. By introducing a superpopulation distribution over the possible (λ_0,λ_1) pairs, one can use these sample path fragments to estimate the superpopulation parameters and the particular (λ_0,λ_1) realizations. Finally, the superpopulation distribution can be used to carry out a complete system performance analysis.

3.2 Compartment Models in Pharmacology

Compartment models offer a broad class of models often used to represent the movement of drugs or pollutants through a system such as the human body. The compartments correspond to pools or tissue groups such as the blood stream or body organs. Stochastic compartment models are often equivalent to an open or closed Jackson network of infinite server queues. The compartmental structure is usually defined by an $n \times n$ transition matrix $\mathbf{P} = (p_{ij})$ with $p_{ij} \geq 0$, $\sum_j p_{ij} \leq 1$ and $p_{i0} = \sum_j p_{ij}$. The structure functions are taken to be $f_{ij}(x_i) = p_{ij}x_i$, $0 \leq j \leq n$. If $\lambda_0 = 0$ and $p_{i0} = 0$, $1 \leq i \leq n$, then the system is closed. Otherwise, it is open. We assume $\mathbf{P}$ is known, but the analysis can be carried out if we replace $\lambda_i p_{ij} x$ by $\lambda_{ij} x$ where $\lambda_{ij} = \lambda_i p_{ij}$ must be estimated.

It has been commented on by Koch-Weser, as quoted by Wagner (1975), that "drug dosages needed for optimal therapeutic effects differ widely among patients. The 'usual dose' of most potent drugs accomplishes little in some persons, causes serious toxicity in others, and is fully satisfactory in a few." This observed variability between patients strongly indicates that the model rate parameters $\underset{\sim}{\lambda} = (\lambda_0,\lambda_1,\ldots,\lambda_n)$ correspondingly exhibit substantial variability. We imagine that each individual draws a $\underset{\sim}{\lambda}$ from a superpopulation distribution, and observe a sample path from each of the compartment processes. One goal is to estimate the parameters of

the superpopulation distribution, because this determines the variation between members of the population. In addition, it is important to estimate each of the individual $\underset{\sim}{\lambda}$ values, since they may be related to patient pathology or classification. Knowledge of the between variability may be used to strengthen estimates of individual $\underset{\sim}{\lambda}$ values.

3.3 Logistic Support for a System Depending Upon Repairable Modules

Successful operation of each of a set of I vehicle systems (e.g., trucks, rental cars, airplanes, or ships) depends upon the operability of important subsystems or modules (tires, engines, communication and navigation subsystems). Suppose modules are failure prone but repairable, and that each module type that is on a vehicle in operation fails independently at an (unknown) Markovian rate λ_{ji}, where j refers to a module of type j, $1 \leq j \leq J$, and i refers to the i^{th} vehicle, $1 \leq i \leq I$. Let the Markovian repair rate for modules of type j be μ_j (unknown), meaning that repair times are independently exponentially distributed. Suppose that, in addition, there are M_j modules of type j on each vehicle, and that the more that are up or operable, the more effective is the overall system operation. In addition there are S_j spare modules of type j in the system, and a repair system that contains R_j service facilities (repair teams, equipment; spare parts and tools are considered separately). Let $X_{ji}(t)$ represent the number of type j units up and either installed on, or awaiting installation on, all systems $1 \leq i \leq I$ at time t. Then under additional stipulations concerning the service protocol, $\{X_{ji}(t),\ t \geq 0\}$ is a multivariate birth and death process. The total number of modules in the system is $\sum_{j=1}^{J} (M_j + S_j)$, the total number of service facilities is $\sum_{j=1}^{J} R_j$, and the relevant problem is to specify near-optimal values of S_j and R_j when data are available to estimate the rates $\underset{\sim}{\lambda}$ when μ_j and a suitable measure of effectiveness are specified.

4 Statistical Inference for Markov Population Processes

We first develop the likelihood function for simple Markov population processes. Later in this section we consider Bayesian inference.

4.1 The Likelihood Function

A Markov population process behaves in a simple fashion. Suppose at time 0 it is in state $\underset{\sim}{s}$. It will remain in state $\underset{\sim}{s}$ for an exponential period of time with rate parameter

$$R(\underset{\sim}{s}) = \lambda_o \sum_{i=1}^{n} f_{oi}(s_i) + \sum_{i=1}^{n} \lambda_i \sum_{j=o}^{n} f_{ij}(s_i) = \lambda_o f_o(\underset{\sim}{s}) + \sum_{i=1}^{n} \lambda_i f_i(s_i)$$

where $f_o(\underset{\sim}{s}) = \sum_{i=1}^{n} f_{oi}(\underset{\sim}{s})$ and $f_i(s_i) = \sum_{j=o}^{r} f_{ij}(s_i)$.

Of course, $R(\underline{s})$ is the rate at which the process leaves state $\underline{s}$.

When this exponential period ends, the process moves to a new state. These transitions are specified for $1 \leq i \leq n$ and $0 \leq j \leq n$ by

$T_{oi}(\underline{s})$ with probability $\lambda_o f_{oi}/R(\underline{s})$,

$T_{ij}(\underline{s})$ with probability $\lambda_i f_{ij}(s_i)/R(\underline{s})$.

The data from the sample path can be reduced to a sequence of states and sojourn times in those states. This can be written as $(\underline{s}^{(1)},S_1),(\underline{s}^{(2)},S_2),\ldots,(\underline{s}^{(m)}, S_m)$ where $\underline{s}^{(t)}$ is the t^{th} state occupied by the process. To remove certain minor difficulties, we will assume that $\underline{s}^{(1)}$ is not random but is deterministic. It is possible to assume that the initial state is stochastic and is chosen by some distribution (usually the equilibrium distribution, if one exists). We do not even assume the existence of an equilibrium and so simplify matters by taking $\underline{s}^{(1)}$ to be deterministic. We have not described the sampling interval [0,T] and prefer to leave it unspecified. T may be deterministic, in which case the number of transitions (m-1) is stochastic. Alternatively, one could observe the process until m-1 transitions have occurred, in which case $T = \sum_{t=1}^{m} S_t$ is stochastic. The likelihood, however, would not differ substantially.

Thus, the sojourn times, $S_1,\ldots,S_m$ contribute a factor of

$$\prod_{t=1}^{m-1} R(\underline{s}^{(t)}) \prod_{t=1}^{m} \exp(-S_t R(\underline{s}^{(t)}))$$

to the likelihood function for given $\underline{\lambda}$. The state transitions also contribute to the likelihood function. If the transition from $\underline{s}^{(k)}$ to $\underline{s}^{(k+1)}$ involves a departure from colony i to colony j, $0 \leq j \leq n$, a term $\lambda_i f_{ij}(s_i^{(k)})/R(\underline{s}^{(k)})$ is multiplied into the likelihood function. A term $\lambda_o f_{oi}(s_i^{(k)})/R(\underline{s}^{(k)})$ is included if an individual arrives to colony i from the outside. We let M_o = the total number of arrivals from the outside, and M_j = the total number of departures from colony j. As a consequence, the likelihood function is proportional to

$$\exp(-\sum_{t=1}^{m} S_t R(\underline{s}^{(t)})) \prod_{k=0}^{n} \lambda_k^{M_k}. \tag{4.1}$$

The quantity

$$\sum_{t=1}^{m} S_t R(\underline{s}^{(t)}) = \sum_{t=1}^{m} S_t \left[\sum_{i=0}^{n} \lambda_i f_i(\underline{s}^{(t)})\right] = \sum_{i=0}^{n} \lambda_i \sum_{t=1}^{m} S_t f_i(\underline{s}^{(t)}) = \sum_{i=0}^{n} \lambda_i W_i$$

where $W_i = \sum_{t=1}^{m} S_t f_i(\underset{\sim}{s}^{(t)})$.

The sufficient statistics are given by $(\underset{\sim}{M}, \underset{\sim}{W})$ where $\underset{\sim}{M} = (M_0,\ldots,M_n)^T$ and $\underset{\sim}{W} = (W_0,\ldots,W_n)^T$. The likelihood function is of the multivariate independent Poisson type:

$$\exp(-\sum_{j=0}^{n} \lambda_j W_j) \prod_{j=0}^{n} \lambda_j^{M_j} \tag{4.2}$$

4.2 Random Parameter Processes

The likelihood given in equation (4.2) is relevant to a single version or realization of a simple Markov population process (SMPP). In many circumstances, it is reasonable to suppose that the rate parameters $\underset{\sim}{\lambda} = (\lambda_0,\ldots,\lambda_n)^T$ occasionally change, for example in response to external events. Assume we are able to observe K such different versions and wish to estimate the individual parameters on the supposition that important between-version variability may exist. The simplest plausible way to proceed is to introduce a superpopulation of parameters $\underset{\sim}{\lambda}$ having density $f(\underset{\sim}{\lambda}|\underset{\sim}{\phi})$ where $\underset{\sim}{\phi}$ denotes a hyperparameter vector. The K observed sample paths are then analyzed as if the parameters of each version were selected at random using the density f. Specifically, let $\underset{\sim}{\lambda}^{(1)},\ldots,\underset{\sim}{\lambda}^{(K)}$ be i.i.d. random vectors with density $f(\underset{\sim}{\lambda}|\underset{\sim}{\phi})$. Here $\underset{\sim}{\lambda}^{(k)}$ corresponds to the rate parameters for the k^{th} observed SMPP, $\{\underset{\sim}{X}_k(t), 0 \le t \le T_k\}$.

The assumption of independent sampling allows one to simply construct an overall likelihood that incorporates the information as well as that in the superpopulation density f.

4.3 Bayesian Inference

Simple Bayesian inference assumes the superpopulation density f to be completely specified. In the case of a superpopulation density $f(\underset{\sim}{\lambda}|\underset{\sim}{\phi})$, the hyperparameters $\underset{\sim}{\phi}$ would be treated as known, presumably by elicitation. Estimation of $\underset{\sim}{\lambda}^{(k)}$ is accomplished by finding the posterior density of $\underset{\sim}{\lambda}^{(k)}$ given $\{\underset{\sim}{X}_k(t), 0 \le t \le T_k\}$ and then computing some appropriate estimator such as the posterior mean vector or the posterior mode vector. In this section, we consider three particular parametric priors.

4.3.1 Conjugate Gamma Prior

The easiest situation occurs when we introduce a multivariate gamma prior distribution with independent marginals. Specifically, we assume

$$f(\underset{\sim}{\lambda} \mid \underset{\sim}{\alpha}) = \prod_{j=0}^{n} \beta_j^{\alpha_j} \lambda_j^{\alpha_j - 1} \exp(-\beta_j \lambda_j) / \Gamma(\alpha_j), \quad \lambda_j \geq 0, \quad 0 \leq j \leq n. \tag{4.3}$$

The hyperparameters are specified by $\underset{\sim}{\alpha} = (\underset{\sim}{\alpha}, \underset{\sim}{\beta})$. For a likelihood function given by (4.2), the posterior distribution of $\underset{\sim}{\lambda}$ given $\{\underset{\sim}{X}(t), 0 \leq t \leq T\}$ is given by

$$f(\underset{\sim}{\lambda} \mid \underset{\sim}{\alpha}, \underset{\sim}{X}) = \prod_{j=0}^{n} (\beta_j + W_j)^{\alpha_j + M_j} \lambda_j^{\alpha_j + M_j - 1} e^{-(\beta_j + W_j)\lambda_j} / \Gamma(\alpha_j + M_j), \quad \lambda_i > 0, \quad 0 \leq j \leq n \tag{4.4}$$

The posterior distribution has independent gamma marginals. It is simple to estimate any of the individual λ_j or some function of them. The posterior mean of λ_j is given by $(\alpha_j + M_j)/(\beta_j + W_j)$, while the posterior mode is given by $(\alpha_j + M_j - 1)/(\beta_j + W_j)$ provided $\alpha_j + M_j \geq 1$. The posterior mode is seen to resemble the mean, and it turns out to be a convenient approximation. We assume that the hyperparameters $\underset{\sim}{\alpha} = (\alpha_0, \ldots, \alpha_n)$ and $\underset{\sim}{\beta} = (\beta_0, \ldots, \beta_n)$ are known, presumably by elicitation and relevant experience.

4.3.2 Multivariate Log-Normal Prior

The previous choice of conjugate prior offers considerable mathematical tractability; however, it does not permit the $\underset{\sim}{\lambda}$ parameters to be correlated. Furthermore, a common choice of prior for univariate Poisson process data in reliability and probabilistic risk assessment studies is a log-normal distribution, see Rasmussen (1975), Hill, Heger and Koen (1984), or Kaplan (1983). We introduce $\epsilon_i = \log \lambda_i$ and let $\underset{\sim}{\epsilon} = (\epsilon_0, \ldots, \epsilon_n)^T \sim N(\underset{\sim}{\beta}, \Sigma)$ where $\underset{\sim}{\beta}$ is an n+1 dimensional mean vector and Σ is an (n+1) x (n+1) positive definite covariance matrix. This formulation provides both log-normal marginals and the possibility of correlated components in a familiar fashion.

The log-posterior distribution is obtained from the prior and equation (4.2) and is given by

$$\log f(\underset{\sim}{\lambda} \mid \underset{\sim}{\alpha}, \underset{\sim}{X}) = \log C - (\underset{\sim}{\epsilon} - \underset{\sim}{\mu})^T \Sigma^{-1} (\underset{\sim}{\epsilon} - \underset{\sim}{\mu})/2 - \underset{\sim}{\lambda}^T \cdot \underset{\sim}{W} + \underset{\sim}{\epsilon}^T \underset{\sim}{M}$$

where C is a normalization constant and $\underset{\sim}{W} = (W_0, \ldots, W_n)^T$ and $\underset{\sim}{M} = (M_0, \ldots, M_n)^T$

It follows that $\log f(\underset{\sim}{\lambda} \mid \underset{\sim}{\alpha}, \underset{\sim}{X})$ is given, up to constants, by

$$\log f(\underset{\sim}{\lambda} \mid \underset{\sim}{\alpha}, \underset{\sim}{X}) = -(\underset{\sim}{\epsilon} - \underset{\sim}{\mu})^T \Sigma^{-1} (\underset{\sim}{\epsilon} - \underset{\sim}{\mu}) - \underset{\sim}{\lambda}^T \underset{\sim}{W} + \underset{\sim}{\epsilon}^T \underset{\sim}{M}. \tag{4.6}$$

One can consider finding the posterior mode by differentiating (4.6) with respect to $\underset{\sim}{\epsilon}$. The first derivative of (4.6) with respect to $\underset{\sim}{\epsilon}$ is

$$-\Sigma^{-1}(\underset{\sim}{\epsilon}-\mu) + \underset{\sim}{M} - \mathbf{D}\underset{\sim}{1} \tag{4.7}$$

where $\mathbf{D}$ is a diagonal matrix with entries $W_j \exp(\epsilon_j)$, and $\underset{\sim}{1}$ is a column vector of ones.

The second derivative of (4.6) with respect to $\underset{\sim}{\epsilon}$ is given by

$$-\Sigma^{-1} - \mathbf{D} \tag{4.8}$$

The matrix $-(\Sigma^{-1} + \mathbf{D})$ is clearly negative definite, thus one can find the unique posterior mode by solving

$$\underset{\sim}{0} = -\Sigma^{-1}(\underset{\sim}{\epsilon}-\mu) + \underset{\sim}{M} - \mathbf{D}\underset{\sim}{1}. \tag{4.9}$$

Equation (4.9) cannot be solved in closed form, but a numerical solution may always be obtained by a Newton-Raphson iteration approach. As a starting solution, one might use the vector of logarithms of the raw rates, i.e., $\epsilon_j^{(o)} = \log(M_j/W_j)$. In case small, occasionally zero, M_j values occur, one might replace M_j by $M_j+0.5$. The Newton-Raphson will take the form

$$\underset{\sim}{\epsilon}^{(\ell+1)} = \underset{\sim}{\epsilon}^{(\ell)} - (\Sigma^{-1}+\mathbf{D})^{-1}\left[\Sigma^{-1}(\underset{\sim}{\epsilon}^{(\ell)} - \mu - \underset{\sim}{M} + \mathbf{D}\underset{\sim}{1})\right]. \tag{4.10}$$

If the suggested starting value $\epsilon_j^{(o)} = \log(M_j/W_j)$ is used, the first iteration leads to

$$\underset{\sim}{\epsilon}^{(1)} = \underset{\sim}{\epsilon}^{(o)} - (\Sigma^{-1}+\mathbf{D})^{-1}\Sigma^{-1}(\underset{\sim}{\epsilon}^{(o)}-\mu). \tag{4.11}$$

This improved estimate of $\underset{\sim}{\epsilon}$ is only a first approximation to the true solution, but it has a familiar form and intuitive content, namely a weighted combination of the raw rates and the prior mean. As the variability of the superpopulation decreases, more weight is put on the superpopulation mean, μ. Both (4.10) and (4.11) exhibit the tendency for linear shrinkage of the raw estimate vector toward μ, irrespective of the relation of the log observed rates to the superpopulation center. Such linear shrinkage behavior also characterizes estimates obtained from the conjugate gamma superpopulations. It appears to be inappropriate for highly discrepant observations that may occur. Some observed rates may actually choose not to conform to the

relatively short-tailed gamma and log-normal distributions, even if the parameters of the latter are obtained from expert judgement. Such outliers may be the consequence of recording errors or they may be the manifestation of unsuspected influences.

To obtain information about the joint or marginal posterior for ϵ_j and hence for $\lambda_j = \exp(\epsilon_j)$, it is necessary to resort to numerical integration. It has been found that an initial Laplace's method approximation, for which one should consult Mosteller and Wallace (1964) or Kadane and Tierney (1984), followed by a few-point Gauss-Hermite integration, can be quite effective; see Gaver (1985) for a discussion of the univariate simple Poisson case. By this approach, one can assess the sampling error of the point estimate achieved from the model estimate suggested above. One can also compute alternative estimators such as the perennial favorite, the posterior mean, or a weighted version thereof.

4.3.3 Multivariate Log-Student t Prior

In order to recognize the possible existence of rates of greater discrepancy than admitted by a gamma or log-normal superpopulation, we introduce the multivariate log-Student distribution. This superpopulation has longer tails than the normal, and hence should yield interesting estimates for the rates. It can be obtained by scale mixing the multivariate log-normal. A formula for the multivariate t appropriate here is

$$f(\underset{\sim}{\epsilon} \mid \underset{\sim}{\mu}, \underset{\sim}{\Delta}, k) = C_{n+1} (\mathrm{Det}(\underset{\sim}{\Delta})^{-1/2} (1+Q(\underset{\sim}{\epsilon}))^{(n+1+k)/2} \tag{4.12}$$

where $Q = Q(\underset{\sim}{\epsilon}) = (\underset{\sim}{\epsilon}-\underset{\sim}{\mu})^T \underset{\sim}{\Delta}^{-1} (\underset{\sim}{\epsilon}-\underset{\sim}{\mu})/k$, $\underset{\sim}{\Delta}$ is a covariance matrix, k is a degree of freedom parameter, and C_{n+1} is a normalization constant. See, for example, Mardia, *et al* (1979), page 57 and associated references therein.

In order to determine the modes of the posterior distribution we proceed as before by examining the log-posterior for $\underset{\sim}{\lambda}$ or equivalently for $\underset{\sim}{\epsilon}$. By omitting irrelevant constants we find

$$\log f(\underset{\sim}{\epsilon} \mid \underset{\sim}{\phi}, \mathbf{X}) = -(n+1+k)\log(1+Q)/2 - \underset{\sim}{\lambda}^T \underset{\sim}{W} + \underset{\sim}{\epsilon}^T \cdot \underset{\sim}{M}. \tag{4.13}$$

The first derivative of (4.13) with respect to $\underset{\sim}{\epsilon}$ is given by

$$-(n+1+k)\underset{\sim}{\Delta}^{-1}(\underset{\sim}{\epsilon}-\underset{\sim}{\mu})/(2k(1+Q)) - \mathbf{D}\underset{\sim}{1} + \underset{\sim}{M}. \tag{4.14}$$

We define $\Sigma = 2k(1+Q)\Delta/(n+1+k)$ and rewrite (4.14) as

$$-\Sigma^{-1}(\underset{\sim}{\epsilon}-\underset{\sim}{\mu}) - \mathbf{D}\underset{\sim}{1} + \underset{\sim}{M}. \tag{4.15}$$

It should be noted that (4.15) is identical to (4.7) which arises with the log-normal superpopulation except that in (4.15) Σ is a function of $\underset{\sim}{\epsilon}$ through Q.

The second derivative of (4.14) can be written as

$$-\mathbf{D} - \left[\Sigma^{-1} - \Sigma^{-1}(\underset{\sim}{\epsilon}-\underset{\sim}{\mu})(\underset{\sim}{\epsilon}-\underset{\sim}{\mu})^{T}\Sigma^{-1}/(n+1+k)\right]. \tag{4.16}$$

It is now possible for there to be multiple solutions of the likelihood equations

$$\underset{\sim}{Q} = -\Sigma^{-1}(\underset{\sim}{\epsilon}-\underset{\sim}{\mu}) - \mathbf{D}\underset{\sim}{1} + \underset{\sim}{M}. \tag{4.17}$$

and there can be more than one local maximum in the posterior distribution. The use of a posterior mean estimate becomes questionable. It will minimize a mean square error criterion but will tend to give an estimate between the two if such modes exist. An alternative approach is to take a data-based initial estimate of $\underset{\sim}{\epsilon}$, $\epsilon_j^{(o)} = \log(M_j/W_j)$. The $\underset{\sim}{\epsilon}^{(o)}$ is used to compute an initial $\Sigma^{(o)}$. Equation (4.17) is replaced by

$$\underset{\sim}{Q} = -(\Sigma^{(o)})^{-1}(\underset{\sim}{\epsilon}-\underset{\sim}{\mu}) - \mathbf{D}\underset{\sim}{1} + \underset{\sim}{M}. \tag{4.18}$$

which is a particular case of (4.17) and which has a unique solution. This solution often gives useful results especially when dispersion between individual rates is large; see Gaver (1985). The matrix Δ has been weighted by an initial discrepancy factor $2k(1+Q)/(n+1+k)$. When this factor is large, then the shrinkage toward μ is reduced. This is a desired effect for the multivariate log-Student t superpopulation which has heavy tails. Values of $\underset{\sim}{\epsilon}$ far from the mean are possible, consequently in such cases the procedure refuses to "borrow strength" when it is unwarrented. Such a procedure is "discrepancy-tolerant" or robust.

5 Empirical Bayes Methods

In this section, we discuss the empirical Bayes approach to these problems. We consider the previous formulations but assume that the superpopulation hyperparameters are not known. They must be estimated from all the observations.

5.1 Conjugate Gamma Superpopulation

In the previous section, we found that the conjugate gamma prior offered an especially tractable estimation solution. We did, however, assume the hyperparameters were known. Now we consider estimating them directly from the K sample paths. This is done by finding the distribution of a sample path conditional only on the hyperparameters $\underline{\alpha}$ and $\underline{\beta}$, that is with $\underline{\lambda} = (\lambda_0,\ldots,\lambda_n)$ integrated out. This takes on a multivariate independent negative binomial form,

$$f(\underline{M},\underline{W} \mid \underline{\alpha},\underline{\beta}) = \prod_{i=0}^{n} \Gamma(\alpha_i+M_i)\beta_i^{\alpha_i}/(\Gamma(\alpha_i)(\beta_i)^{\alpha_i+W_i}). \tag{5.1}$$

We now compute maximum likelihood estimators of $\underline{\alpha}$ and $\underline{\beta}$ given the data from the K sample paths, $(\underline{M}^{(1)},\underline{W}^{(1)}),\ldots,(\underline{M}^{(K)},\underline{W}^{(K)})$.

The log-likelihood function for $\underline{\alpha}$ and $\underline{\beta}$ is given by

$$\ell(\underline{\alpha},\underline{\beta}) = \sum_{k=1}^{K}\sum_{i=0}^{n} (H(\alpha_i+M_i^{(k)})-H(\alpha_i))+\alpha_i\log\beta_i-(\alpha_i+M_i^{(k)})\log(\beta_i+W_i^{(k)}), \tag{5.2}$$

where $H(x) = \log \Gamma(x)$.
The expression in (5.2) must be maximized over $\underline{\alpha}$ and $\underline{\beta}$.

Let us assume for the moment that $\underline{\alpha}$ is treated as known, hence only $\underline{\beta}$ must be estimated. This is done by solving the likelihood equations.

$$0 = \sum_{k=1}^{K} (\alpha_i/\beta_i - (\alpha_i+M_i^k)/(\beta_i+W_i^k)), \qquad 0 \le i \le n. \tag{5.3}$$

We assume that both $\sum_{k=1}^{K} W_i^k$ and $\sum_{k=1}^{K} M_i^k$ are positive. The case in which both are 0 can be handled separately in a straightforward fashion. It can arise only when there is no activity in a particular colony in all the K sample paths.

Equation (5.3) cannot be solved in closed form; however, it does have a unique solution which can be found numerically. This can be seen by rewriting (5.3) for a particular i as

$$\alpha_i K = \sum_{k=1}^{K} (\alpha_i+M_i^k)(\beta_i/(\beta_i+W_i^k)). \tag{5.4}$$

The right side is monotone increasing in β. It is 0 for $\beta_i = 0$ and increases to $\alpha_i K + \sum_{k=1}^{K} M_i^k > AK$ as β_i approaches infinity, hence there is a unique root.

When $\underline{\alpha}$ must also be estimated, we must introduce a second set of likelihood equations.

$$0 = \sum_{k=1}^{K} (H'(\alpha_j+M_j^k))-H'(\alpha_j) - \sum_{k=1}^{K} \log((\beta_j+W_j^k)/\beta_j), \quad 0 \leq j \leq n. \tag{5.5}$$

Equations (5.3) and (5.5) must be solved simultaneously for $\underline{\alpha}$ and β. This problem is the multivariate version of one discussed by Deely and Lindley (1981). Indeed the multivariate problem separates into K univariate ones.

5.2 The Log-normal and Log-student Superpopulation

In order to estimate the superpopulation parameters (μ,Σ) or (μ,Δ) in the normal and student cases an integrated likelihood must be formed, analogous to that for the gamma model. There is no way of avoiding approximation or numerical integration for trial parameter values, followed by a search of some sort to locate the maximum likelihood estimates $(\hat{\mu},\hat{\Sigma})$ in the normal case, or $(\hat{\mu},\hat{\Sigma},\hat{k})$ in the Student case. Such a program has been carried out and tested in the univariate case, both on simulated and observational data, see Gaver(1985). The integration was conducted by taking a preliminary Laplace method (quadratic approximation to the log-posterior) approach, with correction furnished by Gauss-Hermite integration. Such a procedure appears fairly satisfactory even for the Student superpopulation, although the latter sometimes admits two posterior modes. Such a program becomes far more ambitious in the multivariate situation addressed in this paper, and further approximations may well be required in order to reduce computing. Of course, some assessment of the sampling errors associated with superpopulation estimates will also be desirable. It is likely that bootstrapping will be useful, and some experiments in the univariate Poisson-log-Student case have already shown its potential.

6 Summary and Further Remarks

We have presented here an enhanced version of a quite general familiar and useful stochastic model. The enhancement recognizes between-version variation in process parameters; such may be the result of endogenous influences. We have then addressed the problem of process parameter estimation by characterizing the between variability component with the aid of parametric superpopulations. In particular, it has been shown that the familiar linear shrinkage effects often encountered in Bayes analyses using priors of conjugate form are interestingly modulated when longer-tailed, discrepancy-tolerant priors or superpopulations are introduced.

There are many directions for future work, some of which are currently being taken. Models that permit the between component of variability to arise from a multivariate stochastic process, e.g., multivariate log-Gaussian process rates, are attractive when endogenous influences may occur in a time-series-like manner, as may be true of weather or economic effects. Hyperparameter estimation and model diagnosis again present problems. Such problems promise to require the computer-intensive activity that characterizes much of modern statistics and operations research. It is our hope that the results will, in spirit, resemble the various interdisciplinary statistical efforts of Herbert Solomon, to whom this paper is dedicated.

7 References

Berger, J. O. (1984). "The robust Bayesian viewpoint"; in *Robustness of Bayesian Analyses*, J. B. Kadane (ed.), Elsevier.

Billingsley (1961). *Statistical Inference for Markov Processes.* Univ. of Chicago Press, Chicago, IL.

Box, G.E.P. (1980). "Sampling and Bayes' inference in scientific modelling and robustness." Journal of Royal Statistical Soc., 143, A, pp. 383-430.

Cox, D. R. (1965). "Some problems of statistical analysis connected with congestion" *Procedures of the Symposium on Congestion Theory*, ed. by W. L. Smith and W. E. Wilkinson. Univ. of No. Carolina Monograph Series.

Cox, D. R. and Lewis, P.A.W. (1966). *The Statistical Analysis of Series of Events.* London: Methuen.

Deely, J. J. and Lindley, D. V. (1981). "Bayes Empirical Bayes." JASA, 76, 833-841.

Gaver, D. P. (1985). "Discrepancy-tolerant hierarchical Poisson event-rate analyses." Department of Operations Research, Naval Postgraduate School Technical Report.

Hill, J. R., Heger, A. S. and Koen, B. V. (1984). "The Application of Stein and Related Parametric Empirical Bayes Estimators to the Nuclear Plant Reliability Data System." Austin, Texas: The Univ. of Texas. NUREG/CR-3637.

Kaplan, S. (1983). "On a two-stage Bayesian procedure for determining failure rates from experiential data." IEEE Transactions on Power Apparatus and Systems,

Kelly, F. (1979). Reversibility and Stochastic Networks. Wiley, New York.

Kingman, J.F.C. (1969). "Markov Population Processes," JAP, 6, 1-18.

Mardia, K. V., Kent J. T. and Bibby, J. M. (1979). *Multivariate Analysis.* Academic Press: New York.

Morris C. (1983). "Parameteric empirical Bayes inference: theory and applications," JASA, 78, 47-65.

Morris C. (1983). "Parameteric empirical Bayes inference: theory and applications," JASA, Vol. 78, No. 381 (with discussion), pp. 47-65.

Mosteller, F. and Wallace, D. L. (1964). *Inference and Disputed Authorship: The Federalist.* Addison Wesley: Reading, Mass.

Rasmussen, N. et al (1975). "Reactor Safety Study: An Assessment of Accident Risks in U.S. Commercial Nuclear Power Plants." NUREG-75/014, WASH 1400.

Tierney, L. and Kadane, J. B. (1984). "Accurate approximations for posterior moments and marginals." Tech Report, Dept. of Statistics, Univ. of Minn. and Dept. of Statistics and Social Science, Carnegie-Mellon, Univ. Tech. Report.

Wagner, J. G. (1975) *Fundamentals of Clinical Pharmacokinetics.* Drug Intelligence Publications.

DISTRIBUTION LIST

	NO. OF COPIES
Dr. D. F. Daley Statistics Dept. (I.A.S) Australian National University Canberra A.C.T. 2606 AUSTRALIA	1
Prof. F. J. Anscombe Department of Statistics Yale University, Box 2179 New Haven, CT 06520	1
Dr. David Brillinger Statistics Department University of California Berkeley, CA 94720	1
Dr. R. Gnanadesikan Bell Core 435 South Street Morris Township, NJ 07960	1
Prof. Bernard Harris Dept. of Statistics University of Wisconsin 610 Walnut Street Madison, WI 53706	1
Dr. D. R. Cox Department of Mathematics Imperial College London SW7 ENGLAND	1
Dr. A. J. Lawrance Dept. of Mathematical Statistics University of Birmingham P. O. Box 363 Birmingham B15 2TT ENGLAND	1
Professor W. M. Hinich University of Texas Austin, TX 78712	1
Dr. John Copas Dept. of Mathematical Statistics University of Birmingham P. O. Box 363 Birmingham B15 2TT ENGLAND	1

P. Heidelberger
IBM Research Laboratory
Yorktown Heights
New York, NY 10598 1

Prof. M. Leadbetter
Department of Statistics
University of North Carolina
Chapel Hill, NC 27514 1

Dr. D. L. Iglehart
Dept. of Operations Research
Stanford University
Stanford, CA 94350 1

Dr. D. Vere-Jones
Dept. of Mathematics
Victoria University of Wellington
P. O. Box 196
Wellington
NEW ZEALAND 1

Prof. J. B. Kadane
Dept. of Statistics
Carnegie-Mellon
Pittsburgh, PA 15213 1

Prof. J. Lehoczky
Department of Statistics
Carnegie-Mellon University
Pittsburgh, PA 15213 1

Dr. J. Maar (R51)
National Security Agency
Fort Meade, MD 20755 1

Dr. M. Mazumdar
Dept. of Industrial Engineering
University of Pittsburgh
Pittsburgh, PA 15235 1

Prof. M. Rosenblatt
Department of Mathematics
University of California-San Diego
La Jolla, CA 92093 1

Prof. Rupert G. Miller, Jr.
Statistics Department
Sequoia Hall
Stanford University
Stanford, CA 94305 1

Prof. I. R. Savage
Dept. of Statistics
Yale University
New Haven, CT 06520 — 1

Dr. Paul Shaman
National Science Foundation
Mathematical Sciences Section
1800 G. Street, NW
Washington, D. C. 20550 — 1

Prof. W. R. Schucany
Dept. of Statistics
Southern Methodist University
Dallas, TX 75222 — 1

Prof. D. C. Siegmund
Dept. of Statistics
Sequoia Hall
Stanford University
Stanford, CA 94305 — 1

Prof. H. Solomon
Department of Statistics
Sequoia Hall
Stanford University
Stanford, CA 94305 — 1

Dr. Ed Wegman
Statistics & Probability Program
Code 411(SP)
Office of Naval Research
Arlington, VA 22217 — 1

Dr. P. Welch
IBM Research Laboratory
Yorktown Heights, NY 10598 — 1

Dr. Marvin Moss
Office of Naval Research
Arlington, VA 22217 — 1

Dr. Roy Welsch
Sloan School
M. I. T.
Cambridge, MA 02139 — 1

Pat Welsh
Head, Polar Oceanography Branch
Code 332
Naval Ocean Research & Dev. Activity
NSTL Station, MS 39529 — 1

Dr. Douglas de Priest 1
Statistics & Probability Program
Code 411(SP)
Office of Naval Research
Arlington, VA 22217

Dr. Morris DeGroot 1
Statistics Department
Carnegie-Mellon University
Pittsburgh, PA 15235

Prof. J. R. Thompson 1
Dept. of Mathematical Science
Rice University
Houston, TX 77001

Prof. J. W. Tukey 1
Statistics Department
Princeton University
Princeton, NJ 08540

Dr. Daniel H. Wagner 1
Station Square One
Paoli, PA 19301

Dr. Colin Mallows 1
Bell Telephone Laboratories
Murray Hill, NJ 07974

Dr. D. Pregibon 1
Bell Telephone Laboratories - AT&T
Murray Hill, NJ 07974

Dr. Jon Kettenring 1
Bell Core
435 South Street
Morris Township, NJ 07960

Dr. David L. Wallace 1
Statistics Dept.
University of Chicago
5734 S. University Ave.
Chicago, IL 60637

Dr. F. Mosteller 1
Dept. of Statistics
Harvard University
Cambridge, MA 02138

Dr. S. R. Dalal 1
Bell Laboratories - AT&T
Mountain Avenue
Murray Hill, NJ 07974

Prof. Donald P. Gaver — 20
Code 55Gv
Naval Postgraduate School
Monterey, California 93943-5100

Prof. Patricia Jacobs — 1
Code 55Jc
Naval Postgraduate School
Monterey, California 93943-5100

Dr. Guy Fayolle — 1
I.N.R.I.A.
Dom de Voluceau-Rocquencourt
78150 Le Chesnay Cedex
FRANCE

Dr. M. J. Fischer — 1
Defense Communications Agency
1860 Wiehle Avenue
Reston, VA 22070

Prof. George S. Fishman — 1
Curr. in OR & Systems Analysis
University of North Carolina
Chapel Hill, NC 20742

Prof. Guy Latouche — 1
University Libre Bruxelles
C. P. 212
Blvd De Triomphe
B-1050 Bruxelles
BELGIUM

Library — 4
Code 1424
Naval Postgraduate School
Monterey, CA 93943-5100

Dr. Alan F. Petty — 1
Code 7930
Naval Research Laboratory
Washington, DC 20375

Prof. Bradley Efron — 1
Statistics Department
Sequoia Hall
Stanford University
Stanford, CA 94305

Prof. Carl N. Morris — 1
Dept. of Mathematics
University of Texas
Austin, TX 78712

Dr. John E. Rolph 1
RAND Corporation
1700 Main Street
Santa Monica, CA 90406

Prof. Linda V. Green 1
Graduate School of Business
Columbia University
New York, NY 10027

Dr. David Burman 1
Bell Laboratories - AT&T
Mountain Avenue
Murray Hill, NJ 07974

Dr. Ed Coffman 1
Bell Laboratories - AT&T
Mountain Avenue
Murray Hill, NJ 07974

Prof. William Jewell 1
Operations Research Department
University of California, Berkeley
Berkeley, CA 94720

Dr. Tom A. Louis 1
Biostatics Department
Harvard School of Public Health
677 Huntington Avenue
Boston, MA 02115

Dr. Nan Laird 1
Biostatics Department
Harvard School of Public Health
677 Huntington Avenue
Boston, MA 02115

Dr. Marvin Zelen 1
Biostatics Department
Harvard School of Public Health
677 Huntington Avenue
Boston, MA 02115

Dr. John Orav 1
Biostatics Department
Harvard School of Public Health
677 Huntington Avenue
Boston, MA 02115

Prof. R. Douglas Martin 1
Chairman
Department of Statistics, GN-22
University of Washington
Seattle, WA 98195

Prof. W. Stuetzle 1
Department of Statistics
University of Washington
Seattle, WA 98195

Prof. F. W. Mosteller 1
Department of Statistics
Harvard University
1 Oxford Street
Cambridge, MA 02138

Dr. D. C. Hoaglin 1
Department of Statistics
Harvard University
1 Oxford Street
Cambridge, MA 02138

Prof. N. D. Singpurwalla 1
George Washington University
Washington, D. C. 20052

Center for Naval Analyses 1
2000 Beauregard Street
Alexandria, VA 22311

Prof. H. Chernoff 1
Department of Mathematics
M. I. T.
Cambridge, MA 02139

Dr. T. J. Ott 1
Bell Core
435 South Street
Morris Township, NJ 07960

Alan Weiss 1
AT&T Bell Laboratories
Murray Hill, NJ

Operations Research Center, Room E40-164 1
Massachusetts Institute of Technology
Attn: R. C. Larson and J. F. Shapiro
Cambridge, MA 02139

Research Administration
Code 012
Naval Postgraduate School
Monterey, CA 93943 1